Regression Models Course Project

Executive Summary

This report explores the relationship between a set of variables and miles per gallon (MPG) (outcome). It also tries answers the following two questions:

"Is an automatic or manual transmission better for MPG"

"Quantify the MPG difference between automatic and manual transmissions"

Data Processing

Load the data.

```
data("mtcars")
```

Convert the variable am from numeric to factor "automatic" - "manual".

```
mtcars$am<-factor(mtcars$am,labels=c("automatic","manual"))
```

Exploratory Analysis

Create a boxplot to examine the relationship between mpg and transmission type.

```
plot(mtcars$am,mtcars$mpg,ylab="Miles Per
Galon",xlab="Transmission",main="MPG by Transmission Type")
```

Appendix 1, Graph 1

The boxplot shows better mpg for manual transmission cars, but it examines only the relation between transmission and mpg, assuming all other variables are the same, something that it is not true. So a multivariable regression analysis must be done to compare mpg with other variables that influence mpg.

The next logical step is to examine the weight of the automobiles, which is always highly correlated to the mpg, more wt means lower mpg.

```
cor(mtcars$mpg,mtcars$wt)

## [1] -0.8676594
```

Create a boxplot to examine relationship between transmission and wt.

```
plot(mtcars$am,mtcars$wt,ylab="Weight(lb/1000)",xlab="Transmission",main="Wei
ght by Transmission Type")
```

Appendix 1, Graph 2

The boxplot shows that automatic cars tend to be heavier than manual.

Multivariable Regression Model

From the above, in the regression model wt and of course am are going to be used as predictors.

Use of interaction term or not, do anova test.

```
model1<-lm(mpg~wt+am,data = mtcars)
model<-lm(mpg~wt+am+wt:am,data = mtcars)
anova(model1,model)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + am
## Model 2: mpg ~ wt + am + wt:am
##   Res.Df    RSS Df Sum of Sq     F   Pr(>F)
## 1     29 278.32
## 2     28 188.01  1    90.312 13.45 0.001017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The anova test shows that interacrion term should be included in the linear model.

```
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ wt + am + wt:am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6004 -1.5446 -0.5325  0.9012  6.0909
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31.4161     3.0201  10.402 4.00e-11 ***
## wt           -3.7859     0.7856  -4.819 4.55e-05 ***
## ammanual     14.8784     4.2640   3.489  0.00162 **
## wt:ammanual  -5.2984     1.4447  -3.667  0.00102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.591 on 28 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8151
## F-statistic: 46.57 on 3 and 28 DF,  p-value: 5.209e-11
```

This shows that in adddition to transmission, weight and their interaction term have high relation to explain the variation in mpg. The adjusted R^2 is 82% which means that the model explains 82% of the variation in mpg indicating it is a robust and highly predictive model. All the coefficients in this linear model are significant.

```
library(ggplot2)
pr<-predict(model)
qplot(wt,mpg,data=mtcars,colour=am)+
  geom_line(aes(x=wt,y=pr),data=mtcars)
```

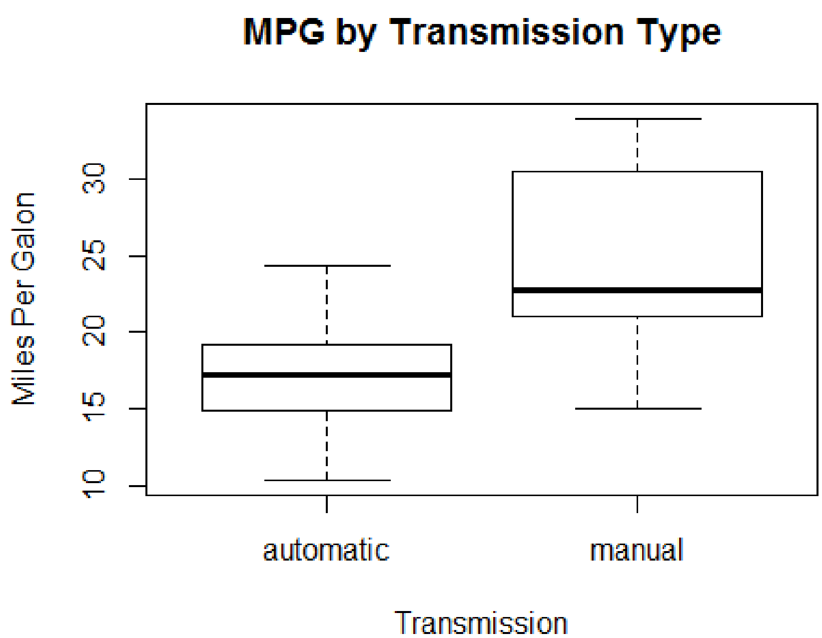
Appendix 1, Graph 3

Results

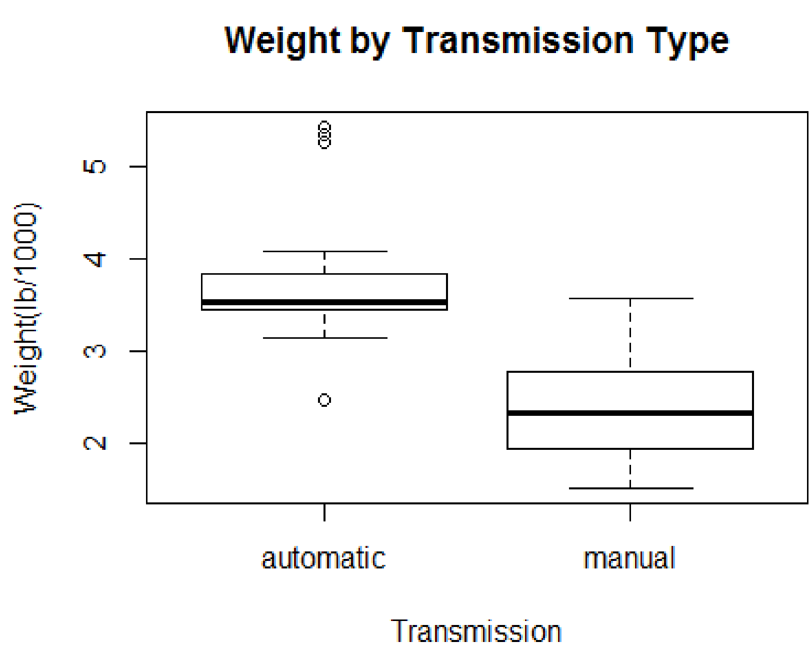
This model explains 82% of the variance in miles per gallon (mpg). It shows why manual transmission cars seems to have more mpg. This is because manual cars in our dataset have less weight than the automatic. The graph 3 shows that the manual cars have bigger intercept but the slope is smaller which means that as weight increases, mpg decreases faster than the automatic cars. So the conclusion for this dataset is that for less heavy cars than 2.75 lb/1000, manual cars have more mpg and for heavier cars automatic are better. For more reliable results we should have more observations for all the weights.

Appendix 1

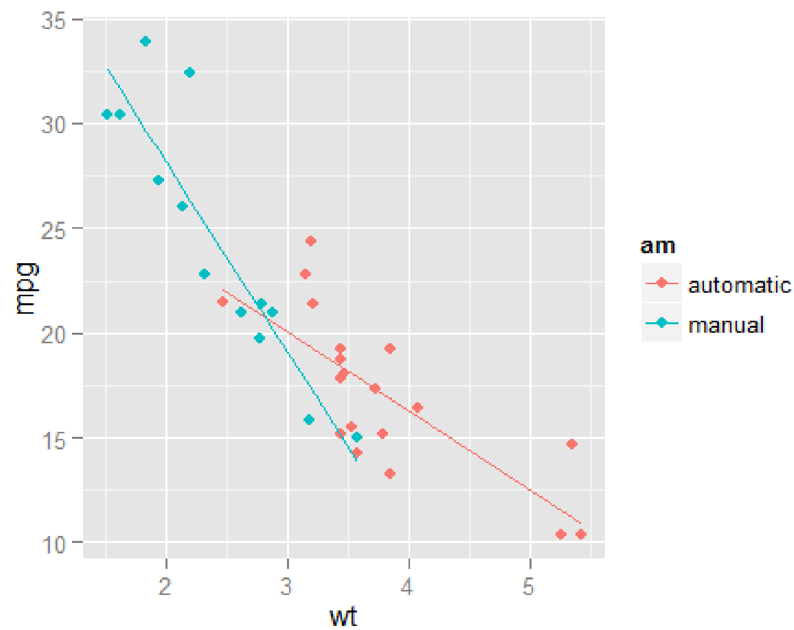
Graph 1



Graph 2



Graph 3



Appendix 2 Residuals and Diagnostics

```
par(mfrow = c(2,2))
plot(model)
```